Simple Strategies for Broadcasting Repository Resources

Carol Minton Morris

NSDL Technical Network Services, Fedora Commons

Cornell University

Tim Cornwell

NSDL Technical Network Services

Cornell University

Abstract

NSDL's data repository for STEM education is designed to provide organized access to digital educational materials through its online portal, NSDL.org[1]. The resources held within the NSDL data repository along with their associated metadata can also be found through partner and external portals, often with high quality, pedagogical contextual information intact[2]. Repositories are not, however, usually described as web broadcast devices for their holdings. Providing multiple contextual views of educational resources where users look for them underscores the idea that digital repositories can be systems for the management, preservation, discovery and reuse of rich resources within a domain that can also be pushed out from a repository into homes and classrooms through multiple channels. This presentation reviews two interrelated methods and usage data that support the concept of "resource broadcasting" from the NSDL data repository as a method that takes advantage of the natural context of resources to encourage their additional use as stand-alone objects outside of specific discipline-oriented portals.

Overview

Common web services such as sitemaps[3] and common blog and wiki technologies have made it simple to send contents from a repository far and wide as web crawlers become tuned to recognize and target certain types of content that is presented in structured formats.[4]

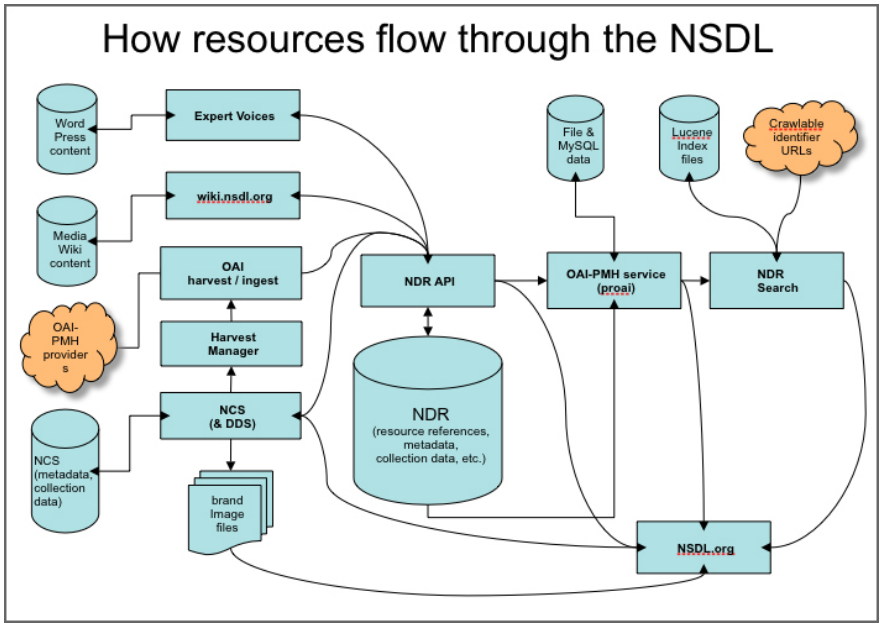

Fig. 1–Identifier URLs (resource references from the NDR) are used to construct on-the-fly NSDL resource "landing pages" that present dynamic user views of NSDL content in search engine indices. Diagram by Tim Cornwell.

For the purposes of this presentation, a resource in the NSDL Data Repository (NDR) is a web page or URL reference. Fig. 1 shows a representation of how most resource references and metadata flow through the NSDL production system. One of the ways that resources are made available for re-use is through Sitemaps.[2] Using NSDL Search and Sitemaps over two million NSDL resources are now available on the Web.

Sitemaps

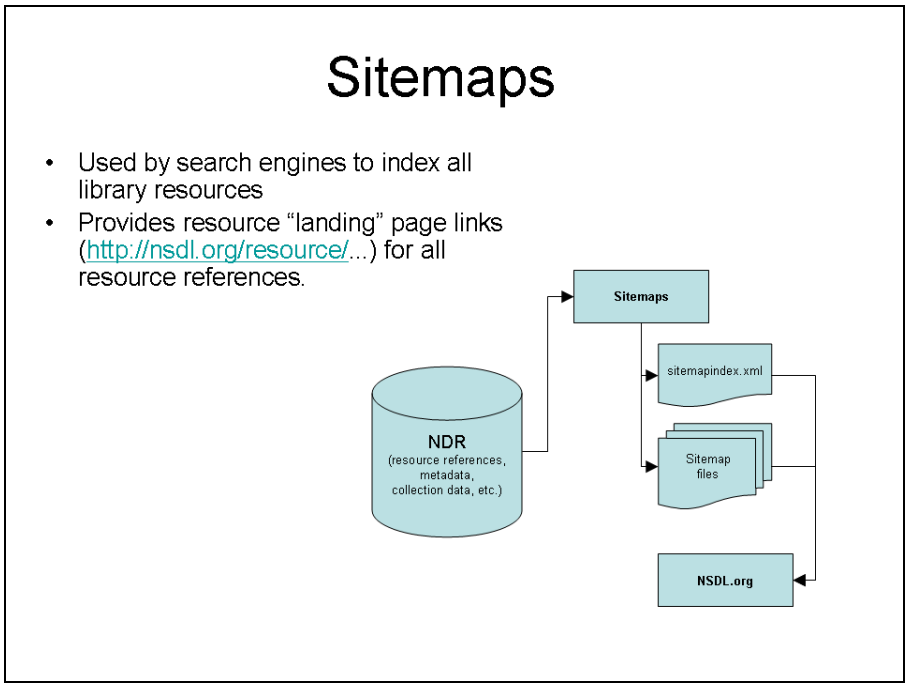

Fig. 2 NSDL Sitemaps currently contain over 2 million individual resource references. Diagram by Tim Cornwell.

The Sitemap protocol[3] provides site managers with a tool for representing a web site's content to crawlers that can be independent of the web site architecture. Using the Sitemap protocol, crawlers like Google can be directed to add appropriate web page content to their indexes—where most teachers, students, instructors and professors find educational resources. Sitemaps have been used most frequently to expose the underlying content of a web site that may not be easily reachable by a site's browse mechanism (if one exists), or that may be ignored by normal crawling methods. Sitemaps are also a means to expose what is sometimes called the "deep web" to outside indexing services. By purposefully allowing all repository content and context to be crawled by major search engine bots, the NSDL Sitemap implementation has exposed individual and collection level resources to anyone using these search engines. Users of Google, Yahoo, MSN, and others can now get search results with direct references into the NSDL's library of STEM resources.

This strategy has resulted in NSDL resources primarily being discovered via the web-based landing pages external to the NSDL domain.

> "The NSDL landing page is the most popular page on NSDL (82,139 page views) so far in 2009 and represents the top way users enter NSDL."[8]

One key to the NSDL's sitemap success has been the internal content pages that the sitemaps refer to. These resource landing pages are generated dynamically, and represent a view of relevant information the library has about a single resource. Because these resource pages have become access points into the library for many library users, they need to be a carefully constructed mix of content and context in order to attain quality rankings by search engines and to provide useful points of entry for users of the NSDL.

The NSDL catalogs high-quality metadata for many of its resources and presents it within the landing pages when available. This allows standard crawling mechanisms to collect page content information and provides a way for future crawlers to potentially improve search accuracy without affecting current crawl methods. The NSDL's resource landing pages provide a "tune-able" representation of the library's resources for consumption by search engines and for presentation to the searching public.

The NSDL released the full set of sitemap files for resources at the beginning of August of 2007. Since that time, trends for access to the NSDL.org site have leaned heavily toward access to the collection through the resource landing pages, contributing 57% of the entry pages to the NSDL in 2008.[7]

Natural Context

Krishna Bharat, Google News Principle Scientist spoke at a Symposium on Computation and Journalism at Georgia Tech on Feb. 22, 2008.[6] He reiterated that new, frequently updated, and well-written content with lots of links to other perspectives would be found and ranked highly by Google News.

NSDL blog and wiki services[5] provide users and resource contributors with facilities that

they can use to call attention to resources within their own repositories, or in the NSDL data repository by:

1. Adding resources to the NSDL repository through blog or wiki services, leveraged by Sitemaps
2. Creating additional regularly crawled natural context around resources in the NSDL repository

By adding new and constantly changing context about repository resources in blog posts and wiki articles search engines "stay tuned" to discover natural links as they recognize URLs that feature "constant context." Stories, images, podcasts, links and snippets about resources that have been added to the repository via NSDL blogs, for example, have been found with high web rankings[2] through interconnected blogs, microblogs, tags, social network spaces and email servers that are cross-referenced in search indices.

One Example: Measuring Focused Web 2.0 Communications

In one 2008 experiment an NSDL partner web site was engaged to test how well contextual content about resources created using NSDL blogs and communication channels affected referrals to the partner web site.

NSDL partners Mimi Recker and Bart Palmer from the Instructional Architect[9] project assisted by observing what happened at their web site when we "stirred the NSDL semantic soup." The goal was to find out what effect pushing communications out from NSDL through interconnected channels using NCore tools in five different ways had on traffic at the Instructional Architect web site.

A short article about Mimi's IA research at Utah State University was posted and reflected out beginning on June 2, 2008 in the following ways:

1. Expert Voices NSDL Highlights > NSDL.org homepage for one week. http://expertvoices.nsdl.org/highlights/2008/06/02/instructional-architect-helping-to-design-the-digital-classroom/
2. Catalog record generated in On Ramp > Published to Whiteboard Report distribution on NSDL.org and through email lists http://onramp.nsdl.org/view/onramp:810
3. Posted in Expert Voices Whiteboard Report Talk Back http://expertvoices.nsdl.org/whiteboardtalkback/2008/06/05/add-your-comments-to-whiteboard-report-135-instructional-architect-higher-ed-critical-thinking-scholarly-wiki-all-around-a-molecule/
4. Sent out through Yahoo! Teachers network
5. Indexed by Google and Yahoo

Notes on Google rankings: Content was ranked #1 on the terms (title), "Instructional Architect is Helping to Design the Digital Classroom"; #4 on the terms "Instructional Architect"; #8 (OnRamp record) and #10 (ExpertVoices post) on the terms "designing digital classroom" search.

Notes on Yahoo rankings: This content was ranked #1-10 (links and references to the NSDL content) on the terms (title), "Instructional Architect is Helping to Design the Digital Classroom"; #11 on the terms "Instructional Architect"; #15 on the terms "designing digital classroom."

Instructional Architect traffic and web site interactions were monitored beginning on Monday, June 2, 2008.

On July 10, 2008 Bart Selman reported, “I found an interesting difference in the 15 days preceding and 15 following this 'stirring'. Visits referred from the NSDL domain were up while most everything else was losing steam at the end of the school year.

Our site-wide bounce rate went from 30% to 60% (the number of one-page-visits over the number of total visitors) indicating that people clicked on the link, but many did not do anything more.

Referrals from NSDL and Expert Voices both went up while nearly all other referral sources went down across the two weeks.”

Conclusion

It is unclear how specific communications services interoperate with sitemaps to create increased access to NSDL repository resources. There are, however, effects that have been measured and observed that indicate increased access and awareness as a result of pushing resources out from NSDL. More of these simple, one-off experiments in partnership with NSDL projects using NCore tools[10] will continue to tell the story of the value of an emerging semantic communications network, with NSDL educational technology at the hub.

Acknowledgements

Thanks to NSDL colleagues Mimi Recker, Bart Selman, James Blake, Aaron Birkland Carl Lagoze, Sharon Clark and Kim Lightle. This research is based upon work supported by the National Science Foundation under Grant No. 0840774. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

---

1. Krafft, D., Birkland, A., Cramer, E., NCore: Architecture and Implementation of a Flexible, Collaborative Digital Library. Proceedings of the Joint Conference on Digital Libraries 2008, Pittsburgh, PA. http://arxiv.org/abs/0803.1500
2. Edmondson, B., How to Improve Rankings With Blog Posts. NSDL Wiki 2008. http://wiki.nsdl.org/index.php/How_To_Improve_Rankings_With_Blog_Posts
3. Sitemaps.org web site. http://sitemaps.org/
4. Clark, S., Search Engine Optimization Tips, Tricks and Resources blog post. Search Engine Optimization Strategies blog 2008. http://expertvoices.nsdl.org/techtalk/
5. Minton Morris, C. Cramer, E., Embedding the Managed Repository in National Science Digital Library Semantic Library. Proceedings of the Third International Open Repositories Conference 2008, Southampton, UK. http://pubs.or08.ecs.soton.ac.uk/6/

6. Bharhat, K., Video of Keynote Address. Proceedings of Journalism 3G. February 22-23, 2008, Georgia Institute of Technology, Atlanta, GA. http://www.computational-journalism.com/symposium/videos/
7. Clark, S., NSDL Useage Report October 2008 (Omniture Webmetrics). NSDL On Ramp Repository. http://onramp.nsdl.org/view/onramp:16525
8. Clark, S., NSDL Useage Report January 2009 (Omniture Webmetrics). NSDL On Ramp Repository. http://onramp.nsdl.org/view/onramp:16579
9. Instructional Architect web site: http://ia.usu.edu/
10. NCore web site: http://ncore.nsdl.org